Computational microscopy of the supramolecular organization of the respiratory chain complexes

Arnarez, Clement

CHAPTER IV

Cardiolipins and Cytochrome *c* Oxidase

This chapter is based upon the manuscript:

Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels

C. Arnarez, S.J. Marrink, X. Periole, Scientific Reports 2013, 3, 1263—1270

Abstract

The respiratory chain or oxidative phosphorylation system (OxPhos) generates most of the chemical energy (ATP) used by our cells. The cytochrome c oxidase (CIV) is one of three protein complexes of OxPhos building up a proton gradient across the inner mitochondrial membrane, which is ultimately used by the ATP synthase to produce ATP. We present molecular dynamic simulations of CIV in a mimic of the mitochondrial membrane, and identify precise binding sites of cardiolipin (CL, signature phospholipid of mitochondria) on the protein surface. Two of these CL binding sites reveal pathways linking CLs to the entrance of the D and H proton channels across CIV. CLs being able to carry protons our results strongly support an involvement of CLs in the proton delivery machinery to CIV. The ubiquitous nature of CL interactions with the components of the OxPhos suggests that this delivery mechanism might extend to the other respiratory complexes.

Introduction

Mitochondria are the intra-cellular organelles that produce most of the chemical energy consumed by a cell and are therefore often refered to as the "power plant" of the cell. The energy is produced through the oxidative phosphorylation (OxPhos) system. The OxPhos system is embedded in the mitochondrial inner-membrane and consists of four large membrane protein assemblies, the so-called "respiratory chain complexes" (*i.e.*: NADH quinone oxidoreductase, complex I; cytochrome bc_1, complex III; cytochrome *c* oxidase, complex IV; ATP synthase, complex V) and by some small electron carriers (quinones and cytochrome *c*). Complexes I, III and IV operate a series of electron transfers whose tortuous paths through the respiratory chain system is coupled with the translocation of protons from the inside (matrix) to the mitochondrial intermembrane space (IMS), leading to an electro-chemical gradient ultimately utilized by the ATP synthase to transform ADP into ATP.

Cardiolipin (CL), the signature phospholipid of mitochondrial membranes, has crucial implications in mitochondrial processes, and on the OxPhos in particular [60,61]. CLs are formed by two phosphatidyl-glycerol molecules linked by an additional glycerol moiety and possess an anionic charge (-2 *e*). They are found in several parts of the mitochondria and constitute about 10% of the phospholipids of, for instance, the bovine heart mitochondrial inner membrane [21,65]. CL deficiency correlates with numerous diseases [25] including the Barth syndrome [27] and heart failure [26]. Notably, the formation, stability, and function of individual respiratory chain complexes and of respiratory chain supercomplexes strongly relies on the presence of CLs in the membrane [13,28-30,68]. CLs bound to individual complexes have been reported by several structural studies on various organisms [35,66,102], leading to hypotheses on their possible involvement in a proton uptake pathway [74] and/or in assuring the structural integrity of individual complexes [74,103-105] and supercomplexes (*cf.* Chapter V).

The focus of the present work is on bovine cytochrome *c* oxydase (CIV). It is a transmembrane protein complex that consists of 13 subunits labeled A to M as depicted in Figure 4.1A (the subunit's nomenclature follows the roman number code defined by Kadenbach [106]). Primarily based on purification and delipidation by chromatography and binding affinity assays [102-104,107,108] two classes of CL binding sites have been defined: two sites with high-affinity and one to two additional sites with a lower affinity. The latter have been associated with the structural integrity of CIV and of its dimeric form because these CLs stabilize the subunits G and H (VIa and VIb according to Kadenbach nomenclature [106]), which are mandatory for the formation of the dimer [103,104]. The two CLs binding CIV with high affinity have been associated with the regulation of the electron activity of the enzyme [107,108].

It is only with the most recent crystal structures of CIV [109,110] that CLs were found bound to the protein. In the dimeric form of CIV four CLs co-crystallize with the complex, defining three potential CL binding sites per complex (CL1a,b and CL2 in Fig. 4.1A) after consideration of the dimer's symmetry. The intuitive assignment of the CL binding sites based on their location on the complex was in overall agreement with the in-depth photolabeling study of Sedlak *et al.* [105]. The low-affinity CL binding sites are located at the interface of the dimer and interact with subunits G (CL1a/b in Fig. 4.1A), while a high affinity binding site was found on

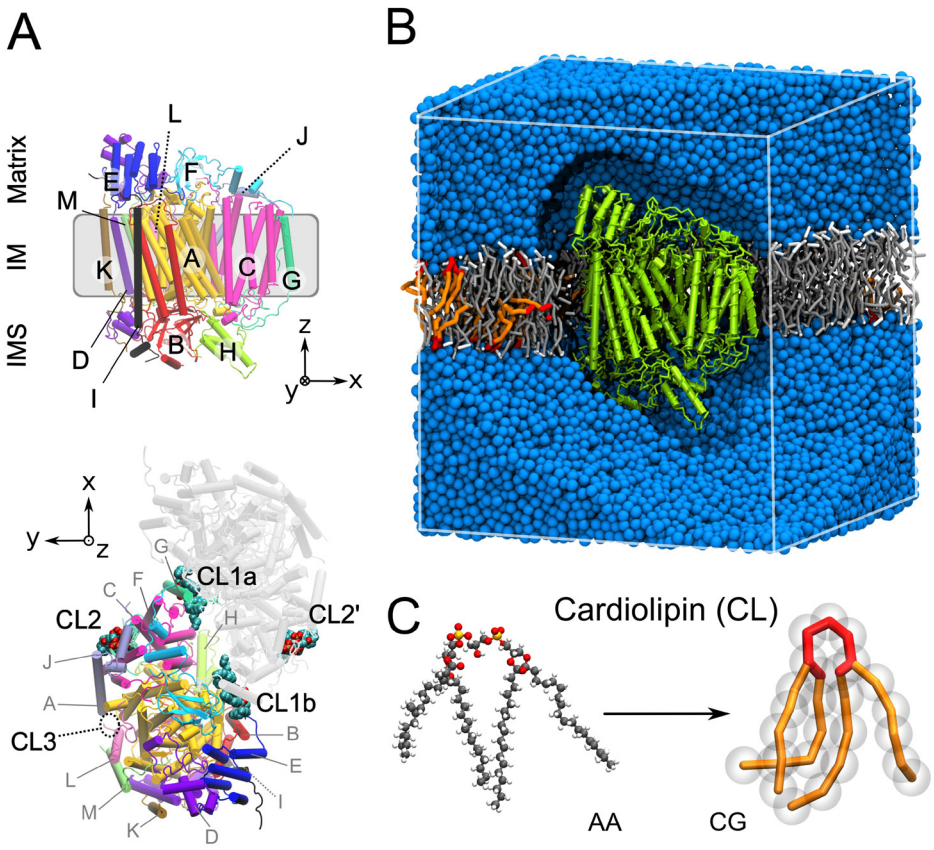

Figure 4.1 | Structural characteristics of bovine heart cytochrome ***c*** oxidase (CIV). **A**) Structure of CIV with its thirteen subunits (A-M) shown in a cartoon representation with a chain-based color-code for one monomer. The subunit's nomenclature is following the roman number code defined and used in others studies [106]. We kept on the letter code to avoid confusion with the site names. In the bottom view of the dimer structure as found in the X-ray structure CLs are depicted [110]. There are three CL binding sites per monomer (CL1a, CL1b and CL2), with one facing the matrix side (CL2) and two facing the intermembrane space (IMS; CL1a and CL1b). CL3 indicates the location of an additional binding site suggested by photolabeling experiments [105]. **B**) Simulation box for CIV system, with the protein shown in green, POPC molecules in gray/white, CLs in red/orange and the aqueous phase in blue. **C**) Structure of a cardiolipin molecule in an atomistic (AA) and a coarse-grain (CG) representation.

a more membrane exposed protein surface interacting with subunit J (CL2 in Fig. 4.1A). There is however a CL binding site identified by Sedlak *et al.* [105] that is missing in the crystal structure (CL3 in Fig. 4.1A).

Here we explore the binding of CL to complex IV in a monomeric form by means of coarse-grain molecular dynamics (CGMD) simulations. MD simulations have been successfully used to explore the lipid/protein interplay [81,111] and more specifically lipid binding to a variety of membrane proteins [57,77-82]. The use of a CG model presents the advantage to investigate the lipid binding to proteins on much longer time scales than conventional atomistic models and thus allows exploring the processes involved from a more realistic dynamic perspective. We recently applied a similar approach to the study of CL binding sites on the cytochrome bc1 [35] in which case we have characterized some key dynamic features of known CL binding sites and revealed a set of new

membrane-exposed binding sites potentially involved in the formation of supercomplexes (see Chapter III).

The set of CGMD simulations presented here clearly identified the precise positions of the two CL with high-affinity binding sites on CIV [105]. The lack of the low-affinity binding sites expected at the interface of the dimer are not observed in the simulations, which confirms that they are strongly associated with the dimer form of CIV and might only exist with the dimer. Five additional CL binding sites with low-affinity are found and may be easily rationalized in light of the position of other co-crystallized lipids [110] and their common features. Most remarkably two of the binding sites are found located at the matrix entrance of known proton uptake pathways (D and H). We show how the structure of the protein reveals extensions of these pathways that directly link the CLs in these binding sites to the entrance of the pathways. In the context of the ability of CL to trap protons our data strongly support that CL maximizes CIV electron transport activity by providing proton to the uptake pathways.

Methods and analyses specific to the work presented in this chapter

Molecular models

The model of CIV was built based on its experimentally determined structure from bovine heart mitochondria. The Protein Data Bank (PDB) entries 1occ and 2occ [112] were used to built a complete model of CIV. Structures of bovine heart mitochondria CIV available in the PDB contain 13 subunits labeled A to M. Although CIV is found in a dimeric form in the crystals, it was simulated as a monomer, its functional form. Details on the complex and its subunits are given in Figure 4.1A.

In the structure of CIV four CLs co-crystallize with the protein in its dimeric form (Fig. 4.1A). The consideration of the symmetry of the dimer arrangement in the crystals results in three CL binding sites per monomer (Fig. 4.1A). One CL binding site, CL2, is located on a membrane-exposed protein surface with its headgroup facing the matrix side of the inner-membrane. The two other sites, CL1a/b, are located at the interface of the dimer of the complex with their head groups facing the IMS. In contrast to earlier studies [105,110] we differentiate CL1a and CL1b because although the symmetry of the dimer makes them equivalent they do form two dissimilar binding sites on each monomer. We simulate CIV in a monomeric form and since they are exposed to the membrane bulk the CL positions described above were not included in the initial conformation of the system.

System setup

The atomic protein structure described above was converted to a CG model and embedded into a pre-equilibrated POPC/CL membrane patch with a 1:20 CL:POPC ratio. The system was energy minimized and simulated for 10 ns with positional restraints on the backbone beads of the protein to relax the solvent and side-chains before starting production runs. The system shown in Figure 4.1B contains the protein (1,780 residues; 4,117 beads), a POPC bilayer (966 lipids; 12,558 beads) including CLs (48 CLs; 1,296 beads) and the aqueous phase (51,549 water beads and 103 sodium ions). In total, the system contains slightly less than 70,000 CG beads and runs at a speed of ~1.3 μs/day using 120 CPUs with a clock rate of 2.6 GHz.

Simulation details

Three 20 μs simulations starting with randomized initial velocities were initially performed and analyzed to judge the convergence of a few observables. The results presented are extracted from a 100 μs simulation, with the first μs of the simulations excluded, considered as an equilibration

period. The root-mean-square deviation of the protein complex (with respect to the experimental structures) typically reached values of 0.3-0.4 nm indicative of its structural stability. All simulations were performed with the GROMACS simulation package [41], together with the Martini CG model [46,50] and the *ElNeDyn* approach [53] (for more details please refer to Chapter II). Parameters for CL were extracted from the work of Dahlberg *et al.* [89].

Analysis

Binding sites definition, occupancy lengths, and residence times were computed as described in Chapter III.

Potentials of mean force (PMFs)

PMFs were used to quantify the binding affinities of specific lipid molecules with respect to CL binding sites on CIV. PMFs were calculated using the umbrella sampling technique. The system used was composed of a CIV embedded in a patch of ~360 POPC and one molecule of the lipid of interest. We tested the binding of CL, TGL and POPC molecules. For each case we ran simulations with a set of umbrella potentials that covered from the lipid-bound situation to free in the bulk membrane (the profile reaches a plateau value). The umbrella potentials were applied to the distance between the center-of-mass of the lipid head group and of the binding site using a harmonic potential with a 1000 kJ mol^{-1} nm^{-2} force constant. This choice resulted in 20 to 30 umbrella simulations spaced by 0.1 nm. We reduced the positional and rotational motions of the protein by applying a weak position restrain on three backbone beads of the protein. These position restraints were not applied on the z coordinate and used a 200 kJ mol^{-1} nm^{-2} force constant (0.42 kcal mol^{-1} $Å^{-2}$). We applied the distance restraint between the centers-of-masses along the direction normal to the protein surface only; therefore an additional weak position restraint (100 kJ mol^{-1} nm^{-2} force constant) was applied on one bead of the lipid head group (GL5 for CL, GLY for TGL and PO4 for POPC) to restrict its exploration of the direction parallel to the protein surface (perpendicular to the normal of the protein surface). The direction normal to the membrane plane was let free. Each umbrella simulation was run for 250 to 700 ns; the first 50 ns were excluded from the analysis. The weighted histogram analysis method (WHAM) [113,114] was used to combine and unbias the simulations, and to produce the free energy profiles. The Bayesian bootstrapping method was used to estimate the error on the calculation.

Results

Identification of seven CL binding sites on bovine heart cytochrome *c* oxidase

We characterized the CL density around the protein from a 100 μs CGMD simulation of this complex embedded in a POPC/CL (20:1 POPC:CL molar ratio) membrane. The average densities of CLs (Fig. 4.2A) demonstrate the existence of several preferential sites of interaction of CL with the protein. We define CL binding sites as the locations having a CL density more than five times the bulk value. For details about the definition and characterization of these binding sites, see the method section of Chapter III. Accordingly, seven binding sites are found on either leaflet of the membrane and are labeled I to VII (Fig. 4.2A). Sites I-V are on the matrix side of the membrane and sites VI and VII on the side of the IMS. Sites V and VII may be occupied by two CLs and are therefore subdivided into Va/Vb and VIIa/VIIb. A hierarchy of binding strength was observed, starting with many CL binding/unbinding events for some sites (*e.g.* 380 exchanges for site VIIb) but only few exchanges for others (*e.g.* few unbinding events for site I) during the same amount of simulation time. For all sites but site I, binding events by at least 4 CLs were observed, allowing us to extract meaningful statistical averaged occupations

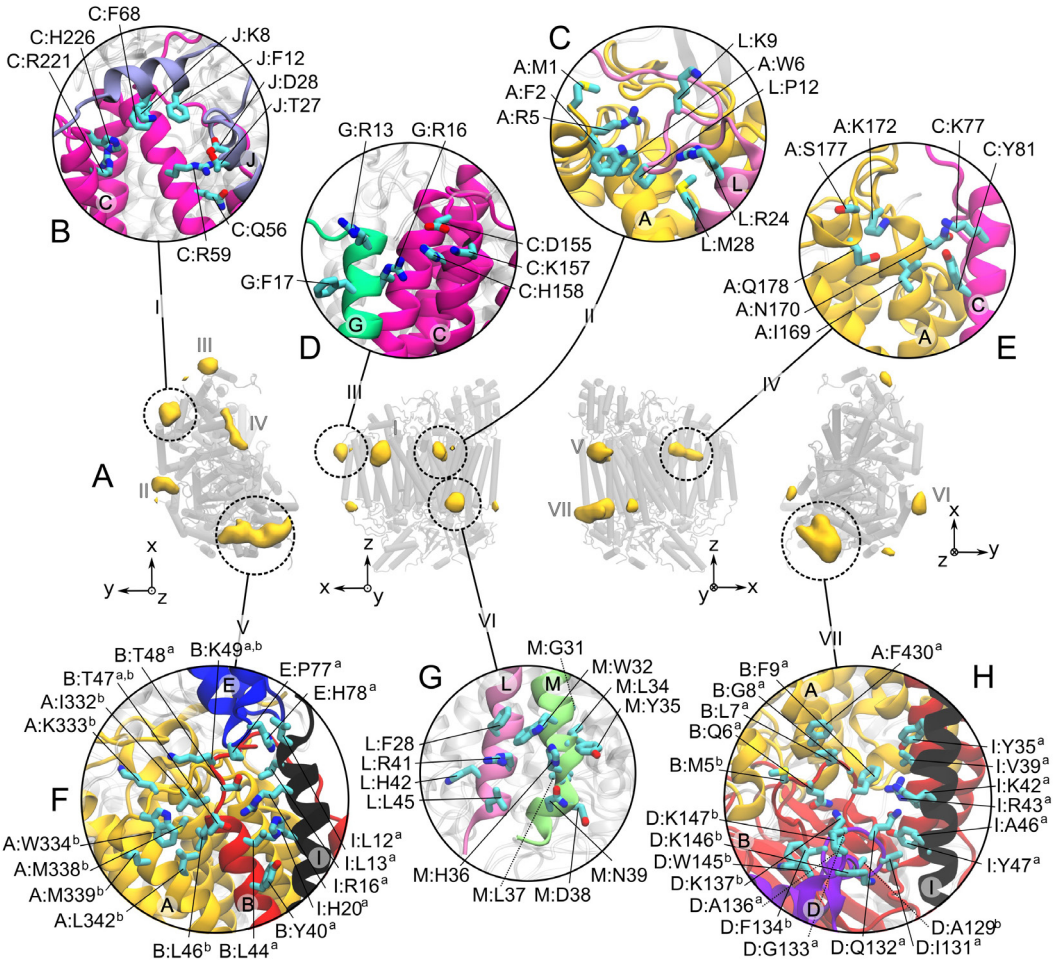

Figure 4.2 | Cardiolipin (CL) binding sites on cytochrome *c* oxidase (CIV). Binding sites are extracted from a 100 μs of CGMD simulation of the complex embedded in a CL/POPC membrane bilayer. **A)** From left to right: matrix, membrane (two orientations) and inter membrane space (IMS) views of CIV with the CL densities shown in yellow volume maps at an isovalue corresponding to at least 5 times the bulk density. The protein is shown as shaded grey cylinders with the CL densities projected onto them. **B-H)** Detailed description of the CL binding sites *i* to VII, respectively. The residues are numbered as follows: "chain:residue^{subsite}". For each site, the subunits involved in the interactions with the CLs are depicted as colored cartoon as in Figure 4.1. The rest of the protein is shown in a transparent gray cartoon.

and CL residence times listed in Table 3.1. The detailed structural characterization of the binding sites is presented in Figure 4.2B-H.

Site I quickly bound a single CL that remained bound during the entire simulation although reorientation of the CL led to multiple short unbinding events (Fig. 4.3). These events complicated the computation of the lifetime estimated at more than 50 μs (Tab. 4.1). In this site the CL mainly interacts with the subunit C and with a couple of residues of subunit J (Fig. 4.2B). Sites II and III are adjacent to site *i* and located at the junction of subunits A and L and subunits C and G, respectively (Fig. 4.2C-D). They both bind one CL at a time and have slightly lower occupan-

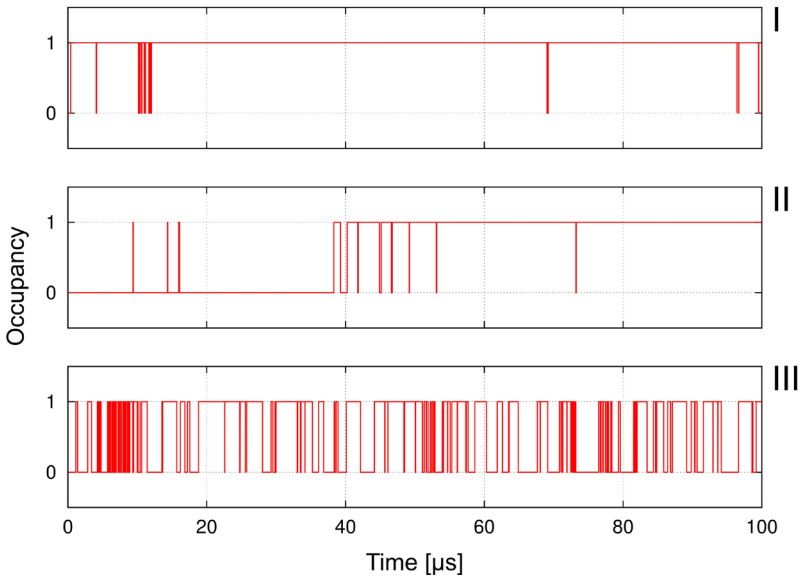

Figure 4.3 | Occupancy of site I to III (top to bottom). Illustration of the hierarchy of CL's binding site strength and dynamics. Sites I and II and strong binder and are only sporadically disrupted. In Site I it is actually the same CL that is bound the all time of simulation. In site II it is a unique CL that binds at t = 40 μs and remains for the rest of the simulation, but three other unsuccessfully attempted to bind earlier in the simulation (t ~ 10–20 μs). Both behaviors contrast with the very dynamic one observed in site III and offering enough statistics to compute the auto-correlation. This analysis was not possible in the cases of site I and site II in which cases the statistics is obviously not sufficient. Instead we qualitatively estimated the lifetime to 50 and 30 μs. Much longer simulations would be required to get more precise values.

Table 4.1 | Occupation (Ξ) and residence time (θ, μs) of cardiolipin binding sites on cytochrome *c* oxidase (CIV). The values are averaged over a 100 μs CGMD simulation. The accuracy of the occupation levels is ± 0.02 at most and of the residence times within ±0.1 μs. See the method section of Chapter III for details about the binding sites definition and characterization. #CLs is the number of different CLs getting in contact with a site and #events the number of binding/unbinding events. The lipids described by Shinzawa-Itoh *et al.* [110] (see Fig. 4.7) at the locations corresponding to the CL binding sites found in the simulations are indicated in the bottom row. The CL binding site predicted by Sedlák *et al.* [105] at the site II is also indicated.

Site	I	II	III	IV	Va	Vb	VI	VIIa	VIIb
Ξ	0.99	0.60	0.57	0.36	0.74	0.68	0.52	0.88	0.50
θ	> 50[a]	> 60[a]	1.0	0.5	0.4	0.6	0.4	10.3	0.2
#CLs	1	4	17	11	20	10	13	4	20
#events	18	11	104	109	235	122	147	41	380
Ref. [105]	–	CL	–	–	–	–	–	–	–
2dyr [110]	CL	TGL2	PE3	PE1	TGL3	TGL3	–	TGL1	TGL1

[a] the strong binding of CLs in sites I and II led to poor binding/unbinding events statistics and prevented us from determining accurate residence time; instead we indicate a rough estimate obtained from the binding behavior (Fig. 4.3).

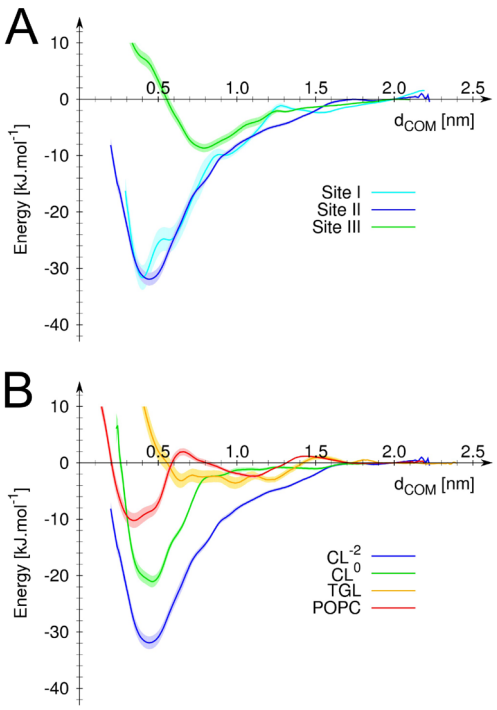

Figure 4.4 | Potential of mean force for binding of various lipids to sites I, II and III. **A**) Comparison of CL's binding strength to site I (cyan), II (blue) and III (green). **B**) Comparison of binding strength of TGL (yellow), POPC (orange) and CL to site II. Two CLs were tested; double charged (–2 *e*) and neutral, blue and green curves, respectively. In both panel the relative free energy of the system is expressed as a function of the distance, d_{CoM}, between the center of mass of the lipid head group and of the binding site as defined in Figure 4.2. The error on the measure (estimated using the Bayesian bootstrapping method) is shown by the shaded area behind the curves

cies (60%) than site I. Although it is difficult to precisely determine the time scale of CLs binding to site II (Fig. 4.3; estimate ~60 μs) it is significantly longer than in site III in which CLs exchange on a time scale of ~1 μs. Site IV is located on the face of CIV that corresponds to the dimer interface in the crystal structure. This site, involving subunits A and C, is occupied only 36% of the time and CLs bind on a sub-microsecond timescale. Site V is located at the other extremity of CIV as compared to site I. CLs in site Va and Vb interact with subunits B, E and I, and A and B, respectively. They have similar occupancies, ~70%, and similar residence times, ~0.5 μs. Site VI is located on the IMS side of CIV. It involves the C-termini of subunits L and M, has a low CL occupancy, 50%, and a small CL residence time, ~0.4 μs. Site VII is located on the same protein surface as site V, but on the opposite leaflet (IMS). CLs in site VIIa and VIIb primarily interact with subunits A, B, D and I on the IMS side of the protein. Site VIIa is significantly more occupied (88%) than site VIIb (50%) and has a relatively long residence time of ~10 μs. The difference between sites VIIa and VIIb might reflect that site VIIa is buried deeper in a cavity formed by helices of subunits B and I, whereas site VIIb is more exposed to the membrane.

We further characterized the binding strength of CL to CIV by determining the potentials of mean force (PMFs) of CL's binding to a selection of the sites (Fig. 4.4). The data clearly indicates that CL's binding to site I and II is stronger than to the other sites. Site III was taken as a reference site because it is relatively well occupied, accessible to the membrane and explored by many CLs so that the statistics obtained is reliable (Tab. 4.1). The data also indicate that CL binds slightly stronger to site II than to site I. The comparison of the PMFs of a CL binding to site II with or without the -2 *e* charge of its head group indicates that although the charge of the head group bears the most significant contribution to the binding, the tails also contribute to CL's binding strength. The contribution from the CL's bulky tail is by itself larger than the binding strength of a POPC molecule.

Residue content of the binding sites

Interesting features emerged from the analysis of the residues in contact with CLs during the 100 μs simulation (Fig. 4.5). First, all the residues that made at least one contact with a CL molecule virtually span the entire transmembrane region of the protein, indicating that the CLs explored the complete

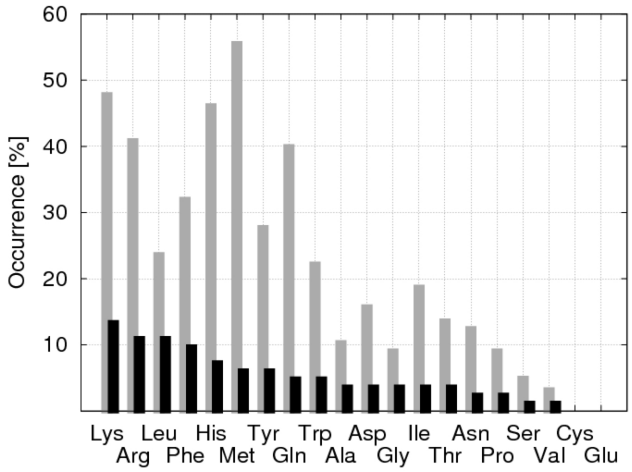

Figure 4.5 | Residue content of the CL binding sites of cytochrome *c* oxidase (CIV). The gray sticks indicate the percentage of each residue type at least once in contact with a CL; *e.g.* present in the section of the CIV accessible to CLs. The black sticks give for each residue type the percentage of its participation to the CL binding sites.

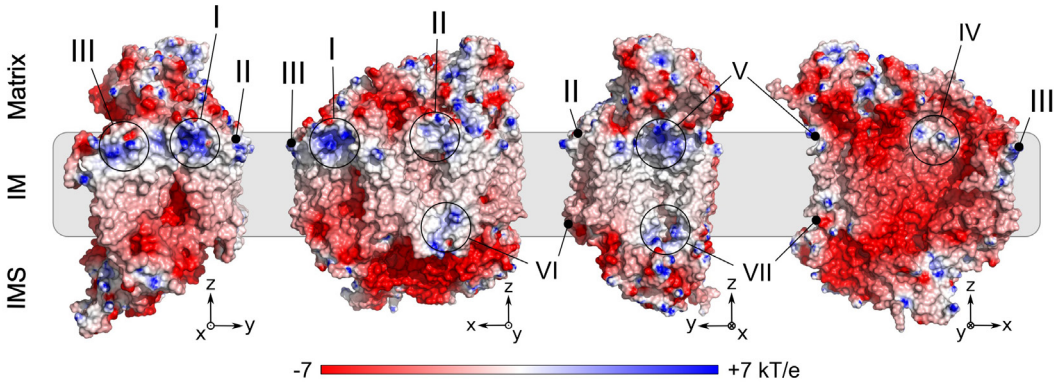

Figure 4.6 | Electrostatic map of the CIV. The electrostatic potential was computed using the *DelPhi* package [115]. Four orientations are shown with the position of the sites defined in this study also indicated. Positive regions are reported in blue, negative in red. Note the strong correlation between regions of the protein surface with a positive electrostatic potential and the location of the CL binding sites (see also CL densities shown in Fig. 4.2). The approximate location of the inner membrane (IM) is indicated by a grey shade. The IM defines the separation between the inter membrane space (IMS) and the matrix compartments.

transmembrane protein surface. As expected, the distribution of contacts as a function of the residue type (grey bars in Fig. 4.5) is in general agreement with the amino acid distribution in integral membrane proteins [94] keeping in mind that a residue would need to be exposed to the membrane to be in contact with a CL. Furthermore, focusing on the composition of the CL binding sites (black bars in Fig. 4.5), the positively charged amino acids (lysine and arginine) were found to account for ~25% of the CL ligands. The large contribution from positively charged residues might be expected since CL carries a -2 *e* charge and is illustrated by the strong correlation between the locations of the CL's binding sites with the positive regions of electrostatic potential on the protein's sur-

face (Fig. 4.6). Lysine is slightly favored over arginine (14 *vs.* 11%), which contrasts with the results obtained previously for CIII [35]. In that case arginine was significantly more represented in the binding sites than lysine, which then represented only ~5% of the CL's ligands against for arginine. Phenylalanine and leucine were found to be a relatively important contribution to the CL binding sites (> 10%) in both CIII and CIV.

Discussion

A thorough analysis of the lipids bound to CIV was presented by Shinzawa-Itoh *et al.* [110] combining a crystal structure of the complex with mass spectroscopy. They discussed up to seven species of lipids per monomer of CIV among which three CLs, three phosphatidylethanolamines (PEs), four phosphatidylglycerols (PGs) and three triglycerides (TGLs). These lipid types are found in the native mitochondrial inner membrane [21] (TGL only recently [110]) and their presence in the crystal is therefore not expected to be an artifact of the crystallization process. Their positions are depicted in Figure 4.7 together with the CL binding sites found in the simulations. The comparison of the two sets of molecules allows to unambiguously identifying the two CL binding sites with high-affinity and involved in CIV's proton transport activity, and suggesting the most likely candidates for the low-affinity binding sites that have been linked to CIV's structural integrity. Our interpretation is in complete agreement with earlier predictions of Robinson and co-workers [102,105,108].

Among the three CL binding sites per monomer found in the dimer structure of CIV [110] (CL1a,b and CL2 in Fig. 4.7), only CL2 is observed in the simulations. It is located at the site I. This site is filled early in our simulation and remains occupied during the entire simulation (Tab. 4.1 and Fig. 4.3), which gives it one of the highest binding affinities observed in the simulations. This result is in agreement with its presence in the crystal

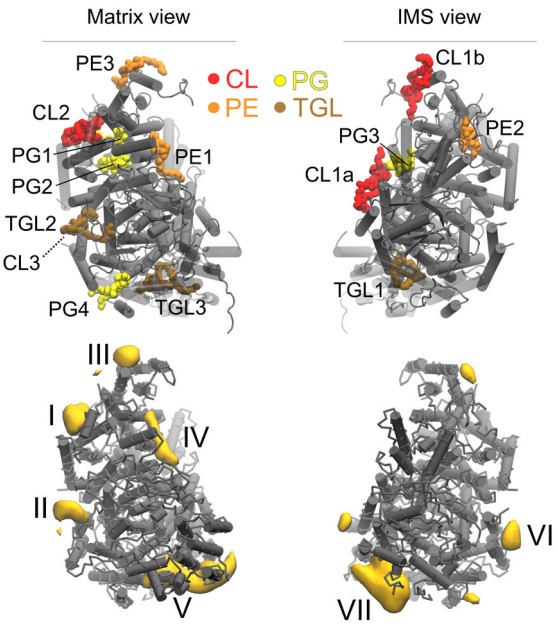

Figure 4.7 | Location of the co-crystallized phospholipid positions and densities computed with our CGMD simulation. The experimental positions (upper part) were extracted from the PBD entry 2dyr [110]. The densities (lower part) are extracted from the 100 μs CGMD simulation. Views from the matrix (left) and the IMS (right) sides of CIV are shown. CL: cardiolipin; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; TGL: triglyceride.

structure and the PMFs described above, and confirms that it corresponds to one of the high-affinity sites as predicted by Sedlák *et al.* [105]. CL1a/b, located at the dimer interface in the crystal structures [109,110], are not stable in our simulations of CIV as a monomer. These two CLs have been suggested to be the ones loosely bound to CIV and strongly connected to the existence of the dimeric form of CIV by their stabilization of subunits G and H [104,105]. Our simulations indicate that the tight association of CIVs in the dimer is a determinant factor for the stability of CL1a and CL1b. To test this hypothesis, additional simulation of the dimeric structure of CIV including the experimental CL1a/CL1b were performed. Both sites were found stable on the μs time scale.

The site II defined by the simulations

(Fig. 4.2 and 4.7) confirms the location of the second high-affinity CL binding site predicted by Sedlák *et al.* [105] using photolabeling and missing in the crystal structures (CL3 in Fig. 4.1). The relatively low occupancy and residence time (Tab. 4.1) might be surprising at first but reflect that it took ~40 μs for a CL to enter the protein cavity forming the binding site (Fig. 3.3). Once the CL made the proper contacts the site is fully occupied and therefore corresponds to a high-affinity binding site. This result is corroborated by the PMF of CL binding to site II, which shows a similar bonding strength to site I (Fig. 4.4A). It is interesting to note that a TGL (TGL2 in Fig. 4.7) occupies the site II in the crystal structure [110]. From our understanding of the system it is extremely unlikely that both CL and TGL molecules occupy site II. The PMFs of the binding of a CL *vs.* a TGL (Fig. 4.4B) demonstrate not only that CL binds much stronger but also that TGL is not stable in site II. Unrestrained simulations confirm this observation: a TGL placed in site II left it quickly and went to mix with the bulk after exploring the surrounding of the protein surface around the site. The PMFs showed that an important part of the CL's binding strength is due to coulombic interactions, which might rationalize the weak binding of TGL; it is a neutral molecule. It is not clear at this point why a TGL molecule is present in the crystal although it was argued not to be an artifact of the purification [110]. One might speculate that the TGL molecule had the capability to replace the CL using the similarity of their bulky tails. This scenario implies however that the CL was first destabilized, which might occur by disruption of the interaction of its head group during the crystallization process. Note the PMFs showed a neutral CL still binds site II stronger than TGL suggesting that TGL's bulky tails are not sufficient to stabilize it.

The remaining CL binding sites (III-VII) found in the simulations are relatively weak binders and unless they are buried within a large protein cavity (sites Va and VIIa) they have a low occupancy. They all correspond to the location of binding of another lipid type in the crystal, which is in line with previous studies showing the conservation of binding sites by different lipid types [66]. We review the several cases observed in CIV and show that it is quite straightforward to rationalize the discrepancies from the similarities of CLs with PGs and TGLs (CL and PG share a negatively charged head group, while CL and TGL bulky tails).

The high degree of conservation of lipid binding sites on CIV from various organisms [66] suggests an important functional role of these bound lipids. CLs have been shown to affect CIV in two ways. First, there is extensive data demonstrating that the presence of two CLs with high-affinity binding to CIV, which we have shown are bound in site I and II, is mandatory for a fully functional CIV [108]. It is however unknown by which mechanism they operate. CIV enzymatic activity is often summarized by its electron transport activity (uptake from cytochrome *c*) but CIV uses as many protons as electrons to transform a dioxygen molecule into two water molecules. CLs have the ability to trap protons [33] and thereby can facilitate proton translocation along the membrane surface [34]. This particular feature of CL has been suggested to be an important aspect of CIII mechanism for proton transfer [74,76,116] and might extend to CIV by providing proton sources to either of the D-, K- or H-pathways [112,117-119]. The proximity of CL3 to the D-pathway entrance has suggested that CL might be acting as a proton antenna to this pathway [105] although only limited structural insight is available since this site is not occupied in the crystal structures [109].

Our data confirm that this ability of CL to carry protons is indeed relevant to sites II but might also be to site Vb. Both sites are located on the matrix side of the IM where the protons are taken up. Site II corresponds to CL3 (Fig. 4.1, 4.7) and is very close to

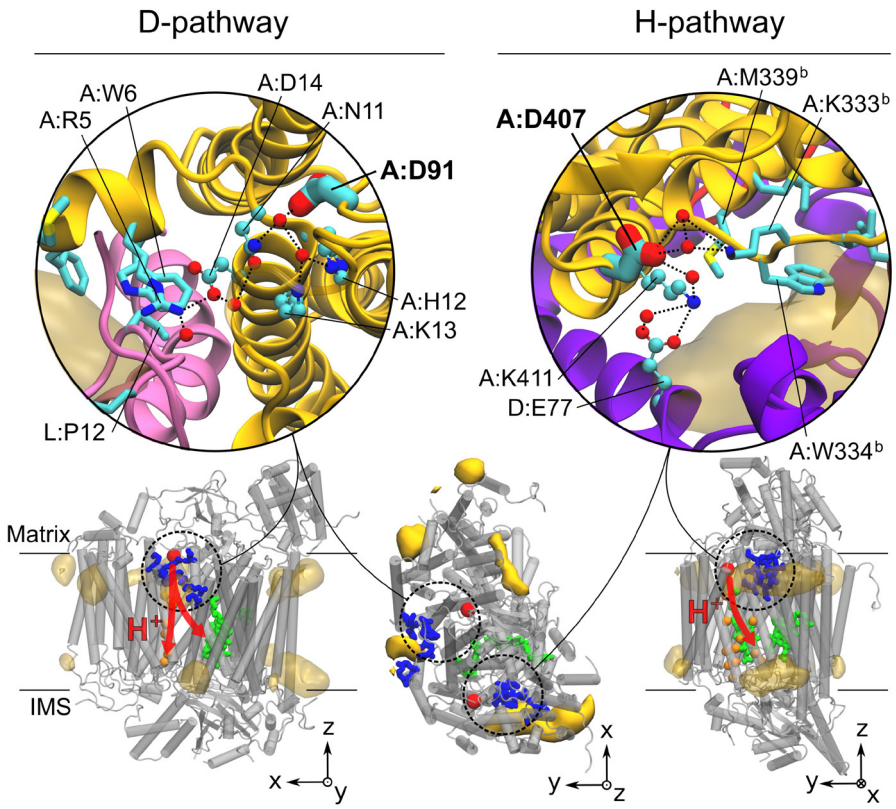

Figure 4.8 | H-bond networks leading from the cardiolipin (CL) binding sites to the entrance of D- and H-pathways, A:D91 and A:D407, respectively. The CL binding sites are shown in stick representation. The yellow surface maps depict the CL densities shown in Figure 4.2. A:D91 and A:D407 are shown in large sticks. The residues and water molecules (red spheres) participating to the networks are shown in a ball-and-stick representation and numbered as in Figure 4.2. The bottom row shows side and top views of CIV with the CL densities, the residues involved in the sites in blue, the residues involved in the transmembrane section of the proton pathways (red arrows) in orange spheres and the heme molecules in green. The large red spheres position the entrances (A:D91 and A:D407) of the pathways.

the D-pathway entrance (~1.1 nm; Fig. 4.8), while site Vb is close to the entrance of the H-pathway (~0.8 nm; Fig. 4.8). Up to now residues A:D91 (A:D132 in *Rodobacter sphaeroides*) and A:D407 define the matrix side entrances to the D- and H-pathways, respectively. Only little is known on the proton path before these residues. A close inspection of the protein structure revealed in both cases the existence of a strong H-bond network starting at the CL binding sites II and Vb and leading to the entrances of the D- and H-pathways, respectively (Fig. 4.8). These networks form clear extensions of the D- and H-pathways towards the exterior of the protein on the matrix side. They directly connect the CL binding sites to A:D91 (A:D132 in *Rodobacter sphaeroides*) and A:D407, respectively, and thereby strongly support the idea that CLs provide a source of protons at the surface of the membrane facilitating CIV electron transport activity.

It is also notable that up to four CLs occupy the protein cavity close to the H-pathway; two lipids in sites Va/b and VIIa/b on the matrix and IMS side of the membrane, respectively. This aggregation of CLs might be relevant to the “dielectric channel” activity of CIV proposed by Rich [120].

The second way by which CLs affect CIV is by providing structural integrity. Two additional CLs, which we propose being CL1a and CL1b in the crystal structures (Fig. 4.1, 4.7), serve the structural integrity of CIV by stabilizing subunits G and H, and thereby its dimeric form [103]. These two CLs have been shown to bind CIV with a low-affinity [105] and are not observed in the simulation of the monomeric form of CIV (Fig. 4.2) but are stable in the dimeric structure. This ability of CLs to stabilize CIV dimers can be extended to high-order oligomers of the respiratory chain complexes as it was reported in the context of the formation of supercomplexes. This behavior is most notable for the formation of supercomplexes $CIII_2$-CIV and $CIII_2$-CIV_2 [28,29]. Our recent CGMD simulations of the CIII [35] suggested that CL binding sites on the surface of the complex might define the location of protein-protein contact with CIV. The CL binding sites found on CIV might share a similar function.

In summary, the CGMD simulations used in this study to investigate the CL binding sites on the respiratory chain complex IV, cytochrome *c* oxidase, provide new and valuable insights on the way CLs participate to the function of proteins. The locations of the CL binding sites that most likely correspond to both the high and low-affinity CL binding to CIV are identified with high level of confidence. They agree with and reconcile all known experimental data. CL binding sites on the surface of the protein are found at the proximity of two of the three known proton-uptake pathways, the D- and H-pathways. Two clear interaction networks connecting the CL binding sites to the entrances of these two pathways are uncovered. To our knowledge they provide the first complete proton pathway between the membrane surface (CL binding sites) and the D- and H-pathways in cytochrome *c* oxidase. This data strongly supports that CLs take active part in the proton delivery mechanism. Given the wide impact of CL in the respiratory chain machinery the role suggested by our data should extend to other components of the respiratory chain. §

CHAPTER V

Cardiolipins and Supercomplexes Formation

This chapter is based upon the manuscript:

Self-assembly of complexes III and IV of the respiratory chain into supercomplexes

C. Arnarez, S.J. Marrink, X. Periole, in preparation

Abstract

Mitochondria produce most of the ATP consumed by the cells through the so-called respiratory chain. It is now well established that the protein complexes forming the respiratory chain assemble into larger structures, the supercomplexes. Cardiolipin (CL), signature phospholipid of mitochondria, composes ~20% of the mitochondrial inner membrane and is crucial for the formation and the functionality of these supercomplexes. Using coarse-grained molecular dynamics we have previously shown that CL forms specific binding sites on the membrane exposed surface of the complex III (CIII) and complex IV (CIV) suggesting an involvement in proton delivery and supercomplex formation. Here we investigate the role of CLs and these bindings sites in the formation and stabilization of protein-protein interfaces by comparing large scale self-assembly simulations of CIIIs and CIVs embedded in bilayer with and without CLs. We observe the progressive formation of supercomplexes in both systems over 20 μs. In the absence of CLs, a significantly larger protein burial is observed due to formation of more CIV/CIV interfaces and an increased number of direct contacts between the complexes. When CLs are present, CLs strongly populate CIII/CIV interfaces, suggesting a potential role in forming and strengthening these interfaces. The slow dynamics of the protein aggregates and the limited number of interfaces formed in the simulations prevents a more quantitative assessment on the relative orientation and preferred stoichiometry of the complexes in supercomplexes. Together, our data show that lipids act as lubricant during the formation of supercomplexes, and CLs at the interface act as traffic control agents steering the complexes together.

Introduction

The respiratory chain, hosted in the inner membrane of the mitochondria, is responsible for the synthesis of most of the ATP used by our cells. Following a complex path, electrons are transported and exchanged by two main ligands — ubiquinone and cytochrome *c* — and three different protein complexes —NADH dehydrogenase, (complex I; CI), cytochrome bc_1 (complex III; CIII) and cytochrome *c* oxidase (complex IV; CIV) — through a set of oxido-reduction reactions. The exchange of these electrons results in a proton gradient across the membrane, ultimately used by the ATP synthase (complex V) to synthesize ATP from ADP.

The respiratory chain electron transfer process has been extensively studied over the last decades, and hypotheses have arisen for these complexes to function in close spatial proximity to minimize the distance the electron carriers need to travel to release or fetch electrons. First revealed by mass spectroscopy, and in the last ten years with the progress of 3D imaging techniques (electron spectroscopy principally), large assemblies of the respiratory chain complexes have been evidenced. These assemblies are referred to as "supercomplexes" or "respirasomes". Depending on the organism studied, different types of supercomplexes can be found: supercomplexes involving CI, CIII and CIV for mitochondria extracted from bovine heart [7-10] or potato [2,3], and supercomplexes formed of different stoichiometry of CIII and CIV in yeast [11,12]. In plant chloroplasts, analogues of these supercomplexes were found involving photosystem *I* and photosystem II [121]. Highly resolved images in which the relative orientations of each complex can be defined are now available for bovine or yeast and supercomplexes, allowing the building of molecular models. However, the relative orientations of the complexes are quite different between the two models from the two organisms. As it stands, it remains currently unclear which factors determine the structure and stoichiometry of the supercomplexes.

Evidence is mounting that the composition of the mitochondrial membrane plays an active role in steering supercomplex formation. In particular cardiolipin (CL), a phospholipid present in large concentration in the inner membrane of mitochondria, has been shown to play an essential role in the functionality of the supercomplexes [10,28,29]. Our recent simulation studies on both CIII and CIV describe CL binding sites on the membrane-exposed surface of both complexes [35,96]. The fast exchange (~μs) of these CLs with the bulk membrane might explain their absence in the crystal structures. The presence of CL binding sites led us to two hypotheses: either the complexes could be interacting through the CLs located in these binding sites ("bridging" hypothesis) or the complexes could use these CLs on the surface to avoid complexes forming unfavorable interfaces ("locking" hypothesis).

In this work, we study the self-assembly of CIII and CIV into supercomplexes using molecular dynamics (MD) simulations. Due to the size of the protein complexes involved and the slow kinetics of protein-protein association, a coarse-grained (CG) resolution is required [122]. We use the CG Martini force field [46,50] to model a mitochondrial membrane composed of almost 20,000 lipids with 45 protein complexes embedded. In order to describe the differences induced by the presence of CLs in the bilayer, two systems were simulated: a system containing CLs in the lipid bulk, and a system without. The significance of this work is fourfold: first, we provide unprecedented details on the lipid content of the interfaces

between complexes associated into supercomplexes and in particular the involvement of CLs; second, we provide evidence for the "bridging" hypothesis in steering supercomplex formation; third, the observed supercomplex structures might help refine the models derived from electron microscopy data; and fourth, as a pioneering simulation of a multi-protein heterogeneous membrane system, this study provides important data on protein oligomerization in a crowded and mixed membrane environment.

Methods and analyses specific to the work presented in this chapter

Molecular models

The models of CIII and CIV were identical to the ones used in our previous studies (Chapter III [35] and IV [96]). Briefly, the model of CIII was built from a combination of four structures extracted from the Protein Data Bank (PDB codes: 1l0l [84], 1sqb/1sqq [85] and 2a06 [86]). CIII being functional in its dimer form, it was simulated as such; CIII contains 6 hemes and 2 iron-sulfur clusters, but these molecules were not included in this work. The model used for CIV was built from the PDB entries 1occ and 2occ [112]. Even though the crystal of CIV contains 2 complexes, it has been simulated in its hypothesized functional monomeric form. CIV contains two deeply buried hemes molecules, but these latter were not included in this work. CG conformations of CIII and CIV were extracted from 100 μs simulations performed in our recent studies on CL binding sites on the individual complexes (Chapter III [35] and IV [96]). Amongst the 6 CL binding sites observed on each monomer of CIII, 4 (II_{CIII} to V_{CIII}) are located on the surface; these sites are all on the periplasmic side of the bilayer. In the case of CIV 5 CL binding sites were observed on the periplasmic side (I_{CIV} to V_{CIV}) and 2 on the cytoplasmic side (VI_{CIV} and VII_{CIV}).

System set-up

The protein structures were randomly oriented and aligned on a grid (Fig. 5.1). For each system, 9 CIII and 27 CIV were used. In the system containing CLs a 1:15 CL:POPC ratio was used. In the case of the system without CLs in the bulk only 2 CLs embedded in the sites Ia and Ib of CIII [35] were conserved. The other CLs were converted to POPC molecules. All systems were energy minimized and simulated for 10 ns with a 10 fs time step to relax the supplementary patches of lipids and solvent before starting production runs. The system with CLs shown in Figure 5.1 contains 9 CIII and 27 CIV (4,230 and 1,780 residues for each complex respectively; ~195,000 beads), a POPC bilayer (17,462 lipids; ~227,000 beads) including CLs (1,175 CLs; ~32,000 beads) and the aqueous phase (~1,116,000water beads and ~2,600 sodium ions). In total, the system contains slightly less than ~1,400,000 CG beads and ran at a speed of ~270 ns/day using ~2,048 hyperthreaded CPUs with a clock rate of 4.7 GHz.

Buried surface and number of contacts

Buried protein surfaces were used to probe the protein-protein contacts. Buried surfaces or burials ($\bar{a}_{AB} = a_A + a_B - a_{AB}$ where $\bar{a}_{ij}$ is the buried surface between complexes i and j, a_i the accessible surface of the complex i, a_{ij} the accessible surface of complexes i and j taken together) were computed with the *g_sas* tool available in the GROMACS suite, which uses the double cubic lattice method [123]. The probe radius was fixed to 0.26 nm, which corresponds to the van der Waals radius of a Martini bead. Alternatively, the number of contacts between complexes was used as a metric for protein-protein interactions. It is a considerably less computationally demanding approach and reports on basically the same observable. In addition it allows a precise quantification of different components for protein-lipid interactions. Protein-protein and protein-lipids

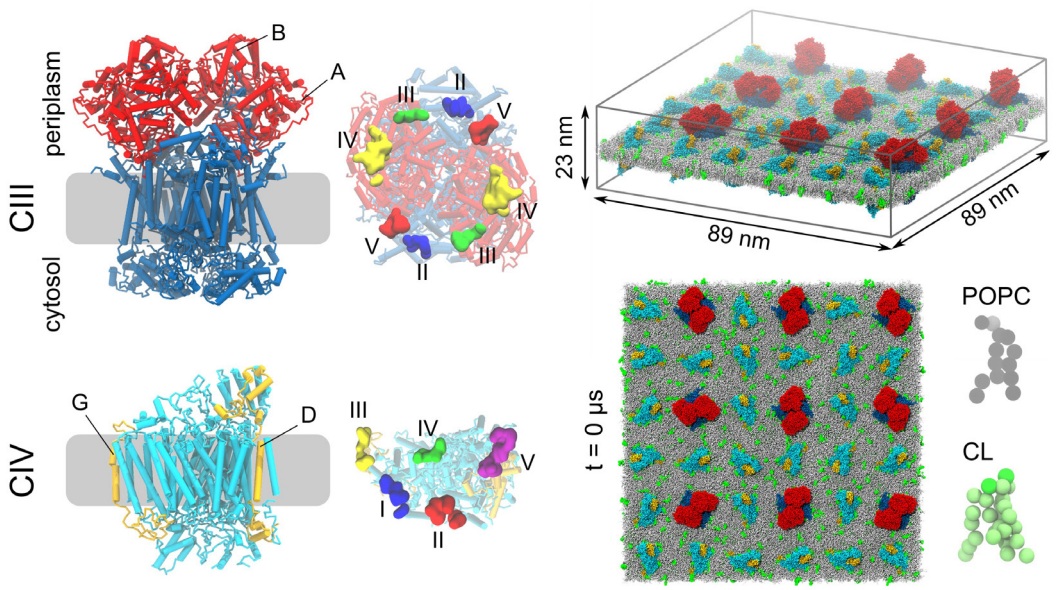

Figure 5.1 | Bovine cytochrome bc_1 (CIII) and cytochrome *c* oxidase (CIV) embedded in a cardiolipin (CL, green) containing POPC (grey) bilayer (right panels). To ease the visualization of the respective orientation of the complexes two subunits of each complex are colored in red (A and B on CIII) or yellow (D and G for CIV). The aqueous phase is omitted. For each complex (left panels), the membrane region is depicted by a gray stripe and a view from the matrix indicates the location of the CL's binding sites on the surface of the proteins [35,96].

contacts were cutoff at 0.8 nm. The differentiation between intramembranous and extramembranous parts of each complex was based on a simple rule: if contacts with lipids are formed at any moment of the previously analyzed 100 µs trajectories, the residue is assumed as being part of the intramembranous section.

Supercomplex architecture

The architecture of the supercomplexes was characterized by the relative orientation of the two complexes following the virtual bond analysis [124] as we did previously in the case of rhodopsin [125]. In the case of a membrane bilayer the motion of the complexes are in two dimensions, reducing the description to one distance and two angles. These were monitored using three anchor points on each complex. For CIII, these three reference points were defined as the center of mass (CoM) of the sites I_{CIII} and I'_{CIII}, the CoM of the four subunits A, A', B and B' and the CoM of site I_{CIII}. For CIV, the CoM of the subunits A, F and C were respectively used. ϕ_1 defines the orientation of CIII relative to CIV — ϕ_3 the relative orientation of CIV to CIII — in the plane of the bilayer. The C2 symmetric axis of CIII dimer was taken into account while determining the CIII/CIV relative orientations.

The CIII/CIV relative orientations corresponding to an ideal overlap between their respective CL binding sites were mapped using a rigid-body docking approach: for each pair of sites on CIII ($site_{CIII}$) and CIV ($site_{CIV}$), CIV was placed such that the distance between $site_{CIII}$ and $site_{CIV}$ CoMs was 2.0 nm. CIV was then moved along the direction defined by the CoM of CIII and of $site_{CIII}$. CIV was then given a full 360° rotation around the CoM of $site_{CIV}$. To increase sampling, the 360° rotation of CIV was performed for each point of a 1.0 nm 3D grid centered on this first point with 0.2 nm spacing. Each con-

formation was associated with the triplet ϕ_1/ϕ_3/number of protein contacts. A triplet was considered favorable when the number of protein contacts (within 0.8 nm) was less than 300, which was the higher value encountered in the supercomplexes formed during self-assembly simulations, and more than 50, which was found to be significant as lower threshold to define an interface.

Results

Formation of supercomplexes

We simulated the time evolution of a model mitochondrial membrane patch containing a set of 9 cytochrome bc_1 dimers (CIII) and 27 cytochrome *c* oxidase monomers (CIV). Two types of lipidic environments were studied: a bilayer containing about 10% of CLs, and a bilayer composed solely of POPC lipids. Initially the protein complexes are distributed on a grid with a random orientation and the CL binding sites populated (see Fig. 5.1). During the subsequent 20 μs simulations we observe the formation of supercomplexes in both systems, as reported in Figure 5.2A by means of representative snapshots from the simulation. The progressive oligomerization results in a final conformation in which most proteins are involved in a supercomplexe. We find abundant formation of interfaces between CIII and CIV, and between CIVs, but only one between CIIIs is observed in the system containing CLs.

To quantify the progressive oligomerization of the complexes, we analyzed the buried surface area, *i.e.* the amount of protein surface area involved in contacts with other proteins. The time evolution of the buried surface is shown in Figure 5.2B. In the system without CLs, the total amount of buried surface is somewhat higher compared to the system containing CLs. In particular, the buried surface area between CIVs is higher, pointing to a preferential interaction between CIVs in the case of the system without CLs. The same conclusion can be drawn from the analysis of the number of interfaces formed during the simulation (Fig. 5.3A), which is substantially higher in case of CIV/CIV in the absence of CLs (16 against 10). On the contrary, just slightly more interfaces are formed between CIII and CIV in the case of the system containing CLs (22 against 19). The maximum number of CIV/CIV interfaces in the system containing CLs formed during the first 14 μs simulation is reached in about 4 μs in the system without CLs.

To shed further light on the nature of the protein/protein contacts, we differentiated between contacts formed in the membrane, and contacts formed in the aqueous phase. This data is shown in Figure 5.3B, normalized per number of interfaces. As expected from the geometry of CIII, containing a large periplasmic domain (*cf.* Fig. 5.1), a significant part of the protein-protein contacts involving CIII are formed between the extramembraneous sites. Interestingly, the absence of CLs results in more contacts inside the membrane for CIII/CIV interfaces whereas the amount of extramembraneous contacts is similar in both systems. This is also apparent from the distributions of buried surfaces associated with each interface formed (data not shown): the interfaces between CIII/CIV are more buried for the system without CLs. In case of CIV/CIV supercomplexes, the presence or absence of CLs does not seem to have a notable effect on the burial per interface. The contacts are shared inside and outside of the membrane in a similar fashion in both systems (Fig. 5.3B).

Even though no unbinding events were observed, the interfaces of the supercomplexes are dynamic. This is shown in Figure 5.3C, where all the contact plots of each individual interface were rescaled to the same time origin. The average number of contacts is reported for both components, inside and outside of the membrane, as well as the associated standard error. The curve of the average follows a positive slope, de-

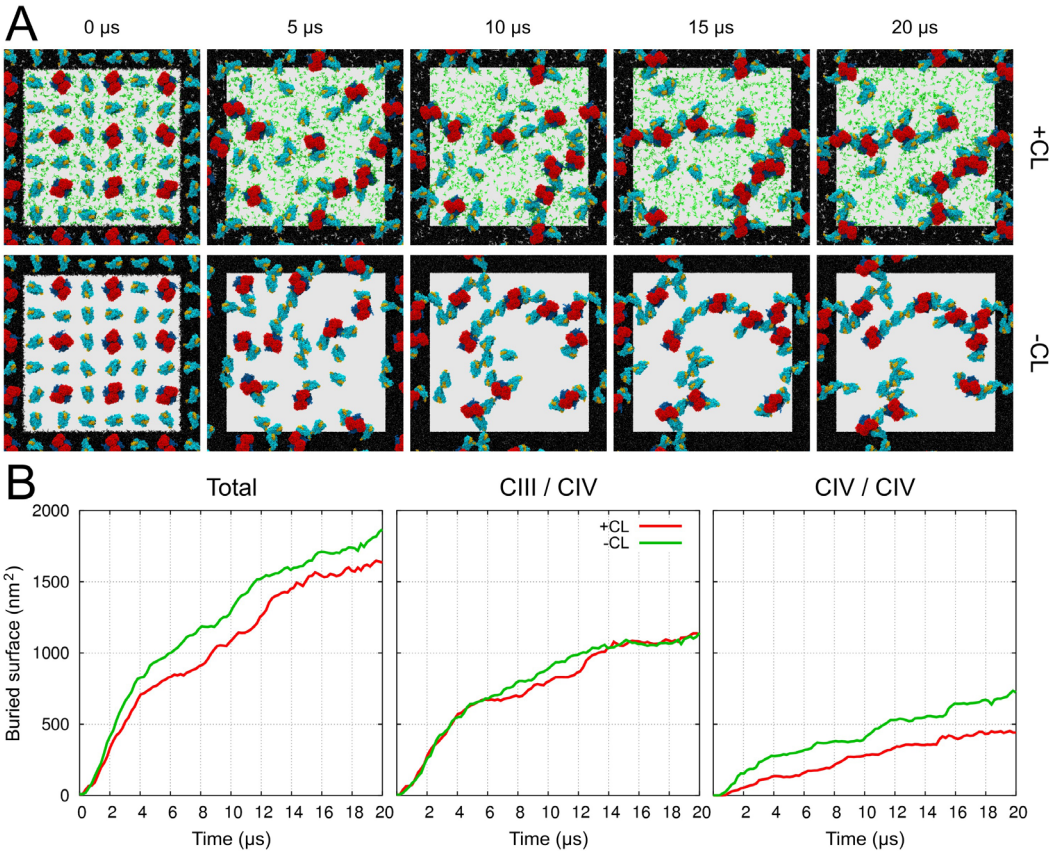

Figure 5.2 | Self-assembly of respiratory chain complexes with (+CL) and without cardiolipins (–CL). **A)** Snapshots of the system viewed from the matrix. The color code is as described in Figure 5.1. The POPC molecules were occulted for clarity. The black frame denotes the periodic image of the unit cell. **B)** Evolution of the proteins burial (left), and the contributions from CIII/CIV (middle) and CIV/CIV interfaces (right).

noting an evolution over time of the interfaces. These rearrangements can be seen as the maturation of the interfaces, slowly evolving to the most stable conformation. This positive slope can be observed for both CIII/CIV and CIV/CIV interfaces, and for each of the two lipidic environments simulated.

The extent of the rearrangements, reported by the time evolution of the buried surface and the relative orientation of the individual complexes, gives further details on the dynamic of the interfaces. Even though the major part of the interaction is formed in the first few microseconds following first contact, significant rearrangements of the interfaces can be observed during the entire simulation as evidenced by a shift of the centers of the buried surface area distributions or changes in the relative orientation of the bound complexes as large as 30°.

In conclusion, we observe the spontaneous formation of supercomplexes involving CIV/CIV and CIII/CIV, but not CIII/CIII, interfaces. Although achieving an equilibrium distribution is out of the question for such a large system (possible requiring millisecond time scales), a number of interesting observations can be made. CLs seem to have a subtle role in steering interface formation, in particular reducing the self-association of

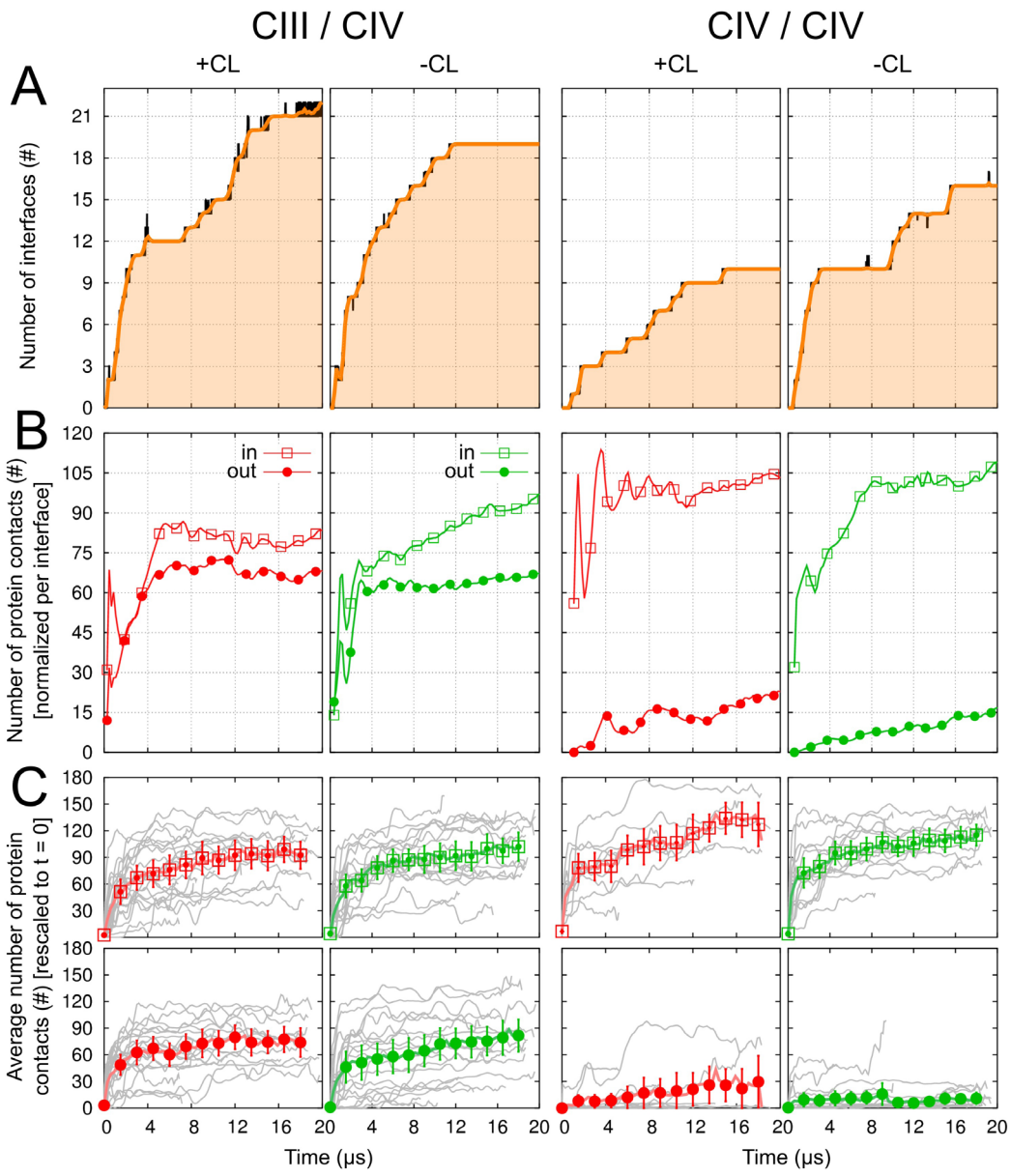

Figure 5.3 | Formation of protein interfaces and protein-protein contacts. **A)** evolution of the number of interfaces between CIII and CIV (left) and CIV and CIV (right) with (+) and without (–) CL. **B)** decomposition of protein contacts, normalized by the number of interfaces as they form in time (represented by the same color code as Figure 5.2): inside (in; square) and outside (out; filled circles) of the bilayer. Values are reported for both CIII/CIV (left) and CIV/CIV (right). **C)** Number of protein contacts as the supercomplexes form. Inner (in) and outer (out) membrane contacts are differentiated. For each interface (light gray) the simulation time has been shifted so that $t = 0$ corresponds to the time of formation of the first contact of the interface formed. The average and standard deviation are given in color.

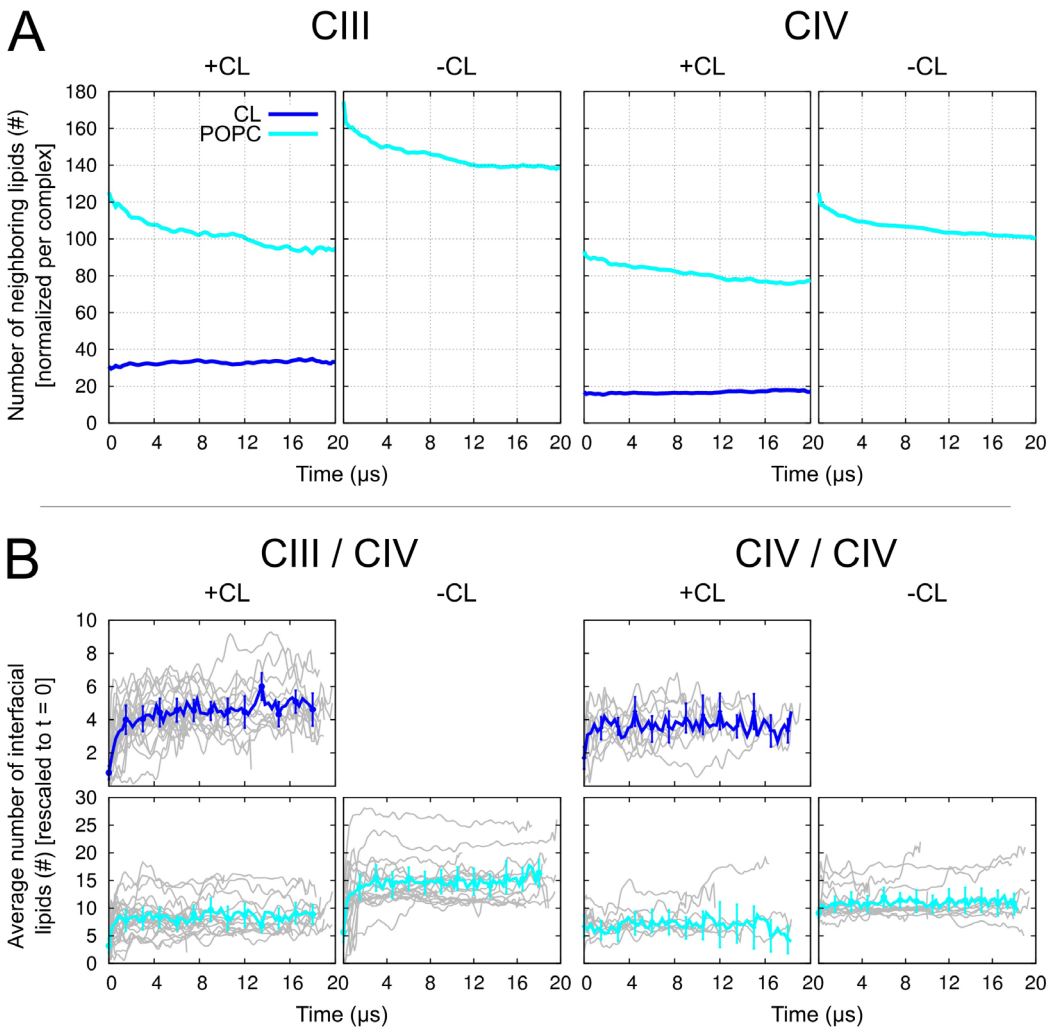

Figure 5.4 | Annular and interfacial lipid content. **A)** Number of annular lipids (< 0.8 nm) per complex (9 $CIII_2$ and 27 CIV) with (+) and without (—) cardiolipins (CL). **B)** Number of interfacial lipids (CLs in blue, POPCs in cyan). For each interface (light gray) the simulation time has been shifted so that $t = 0$ corresponds to the time of formation of the first contact. The average and standard deviation are given in color.

CIV. In the case of CIII/CIV supercomplexes, the presence of CLs lowers the amount of direct protein-protein contacts inside the membrane, pointing to a role of CLs at the CIII/CIV interface. The nature of the interface will be analyzed in more details in the next section.

Phospholipid content at the supercomplex interfaces

We have seen that the lipids present in our systems, and more precisely CLs, can influence the formation of interfaces between complexes in a subtle way. In this section we study this effect in more detail through analysis of the composition of the lipid shell around the proteins. In Figure 5.4A we show the time evolution of the number

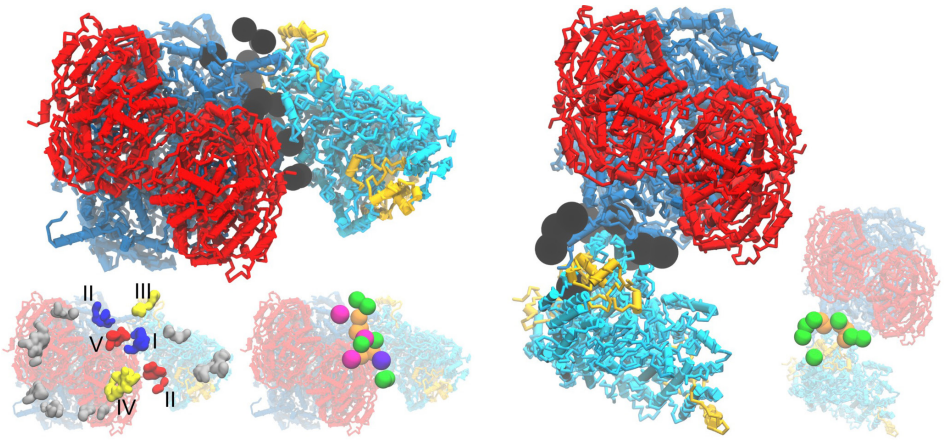

Figure 5.5 | Detailed composition of the interface of the CIII-CIV supercomplexes resembling bovine [9] (left) and yeast [11] (right) EM models. The bovine-like model was obtained with CLs and the yeast-like model without. The CL binding sites involved in the interface are colored as in Figure 5.1, the others are left gray. The lipids found at the interface are shown in black spheres on the large snapshot and in more details in the smaller snapshot: CL and POPC on the periplasmic leaflet are in pink and green and on the cytoplasmic leaflet in purple and yellow, respectively. Only one bead (GL5 bead of CL, PO4 bead of the POPC) is shown for each lipid to ease the visualization.

of lipids in the first solvation shell (lipid head groups residing within nm of the surface of the complexes, see Methods). For both types of membranes a clear decrease in number of POPCs in the first solvation shell is observed as a natural consequence of the oligomerization process. For both systems, a total depletion of ~25 neighboring POPCs for CIIIs, and ~15 lipids for CIVs is observed.

Interestingly, in the system containing CLs, the number of CLs around both types of complexes stays constant. These trends are common to all complex interfaces formed during the simulation. As a consequence of POPC depletion, the average CL:POPC ratio shifts from 1:4 to 1:3 in the vicinity of CIIIs, and from 1:4.5 to 1:4 around CIV*s.* Note, the initial CL:POPC ratio is already drastically higher than the overall 1:15 ratio of the bulk membrane due to the occupied surface exposed CL binding sites as reported earlier (*cf.* Fig. 5.1, Chapter III [35] and IV [96]). The CL binding sites stay occupied during the entire simulation, for both types of complexes, and reproducing the occupancies obtained in our previous studies.

To assess the changes in lipid content right at the protein-protein interfaces, in Figure 5.4B we report the average number of interfacial POPC and CL lipids, with the individual profiles time-shifted to coincide at t = 0, the starting point of complex formation. From Figure 5.4B it becomes apparent that the number of interfacial lipids stays constant once the interfaces are established during the first few μs after the initial contact. On average, the interface between CIII and CIV is composed of 9 ± 1 POPC and 5 ± 0.5 CL lipids in the mixed membrane, and 15 ± 1 POPC lipids in the membrane devoid of CLs. For CIV/CIV, these numbers are 7 ± 1 POPC/4 ± 0.5 CL and 11 ± 1 POPC for +CLs and —CLs membranes, respectively. However, note the large spreads in these numbers for individual complex interfaces (indicated by the grey lines in Fig. 5.4B). The average CL:POPC ratio increases to about 1:2, indicating a strong effective enrichment of CL at the interface Although the total number of lipids at the interface is similar in both membrane environments, the presence at the interface of the much bulkier CL (with four lipid tails) gives rise to more lubricated complex interfaces.

Taken together, our data point to CLs playing an important role in the formation and stabilization of the interfaces. In fact, lipids form an essential part of the interface, with on average 15 lipids present in between the complexes. We observe a strong preference for CLs to be present at the actual protein-protein interfaces, consistent with the increase of number of protein contacts formed inside the membrane when CLs are absent (*cf.* Fig 5.3B). In this latter case, the complexes tend to interact directly (*i.e.* the interface is formed through direct contacts between proteins). Thus CLs may act as a lubricant to steer complex formation.

Interfaces involved in assembled CIII/CIV supercomplexes

We have shown CL lipids are found in between the complexes, lubricating the interface. When absent, tighter interfaces are formed through direct contacts between proteins. Whether the presence or absence of CLs can affect the relative orientation of the complexes will be analyzed and addressed in the following section. Therefore, we calculated the relative orientation of the individual proteins that constitute the supercomplexes. The relative orientations of the proteins are defined by the ϕ_1/ϕ_3 angles (*cf.* Fig. 5.5A for definition), representing the "side" each complex presents to the other interacting complexes. The results of the relative orientation analysis are summarized in Figures 5.5 and 5.6.

From Figure 5.5B, showing the interfacial distribution in case of CIII/CIV supercomplexes, the main conclusion is a lack of convergence. Almost all 22 CIII/CIV interfaces found in the +CLs system as well as the 19 CIII/CIV interfaces of the —CLs system are scattered across the ϕ_1/ϕ_3 landscape. The only exceptions are a relatively well populated area involving CL binding sites III and IV on CIII and Va on CIV in the +CLs system, and a cluster of interfaces around ϕ_1, ϕ_3 (170 ± 100°, 40 ± 100°) in the absence of CL. Comparing the populated interfaces to the interfaces one would predict if CL binding sites were shared between the complexing proteins (Figs. 5.5B, 5.5C), it appears there is a noticeable correlation between the two. However, the same correlation can be observed for both membrane environments, suggesting that there in no evident role of CLs in steering the complex interfaces. The same spread in interfaces is found for CIV/CIV supercomplexes, as reported on Figure 5.5D. In this case, no apparent correlation exists between the interfaces observed and those predicted based on shared CL binding sites (Fig. 5.5D, 5.5E).

A more detailed account of the number of interfaces formed around each complex is presented in Figure 5.6. Here the role of CL becomes more evident. The main differences are found with respect to CIV: i) a significant population of interfaces involving site V of complex IV when CLs are present, and ii) a large population of interfaces around site III in the absence of CLs. Thus, CL bound in site V has a steering role, *i.e.* directing the formation of interfaces around that site. On the contrary, CL in site III appears to have a blocking role, preventing complex formation around that site (which is quite favorable in the absence of CLs).

Although our data show that we are still far from a converged distribution of interfaces, it is interesting to compare the structures formed in our simulations to the experimental models of the CIII/CIV supercomplex. One interface formed during our simulations of the +CLs system is very close to the orientation of the bovine supercomplex as described by Althoff *et al.* [9], indicated by a red mark in Figures 5.5 and 5.6. In the —CLs system, we find an orientation similar to the yeast complex as reported by Heinemeyer *et al.* [11] and indicated by orange marks in these same figures. A graphical visualization of these interfaces is reported in Figure 5.7. Especially noteworthy are the lubricating lipids present in between CIII and CIV, ~8

POPCs and ~4 CLs for the bovine complex, and, ~14 POPCs for the yeast complex that was obtained in the absence of CL.

Summarizing, the experimentally determined models of the CIII/CIV supercomplexes are only a small part of a large ensemble of potential supercomplexes as observed in our simulations. Importantly, we find that the interface of the experimental supercomplex models contain a significant amount of lipids, similar to the other interfaces we observe. Unfortunately, due to our limited statistics, we cannot assess the relative stability of the individual supercomplexes. Likewise, we cannot unambiguously conclude to an effect of CLs in determining the relative orientation of the complexes inside the supercomplexes. The most populated CL binding sites seem to be involved at the interface, but this hypothesis definitively needs more statistics. However, we do see potential roles for CL in blocking as well as supporting complex formation around certain CL binding sites, in particular for CIV.

Discussion

In this work, we have simulated the free diffusion and association of a set of protein complexes from the respiratory chain into their expected multi-component structures, the so-called supercomplexes. Despite the simplification brought by the removal of some degrees of freedom in our CG approach, the systems simulated are extremely large and could be studied at only a fraction of the biological time scales. Therefore, our 20 μs trajectories unravel only parts of the effective mechanism of supercomplex formation, and an equilibrium state has not been reached. In addition, the accuracy to predict protein-lipid and protein-protein interactions is potentially limited by the use of a CG model. Importantly, the Martini model has been shown to reproduce experimental lipid binding sites on an increasing number of membrane proteins [35,96,97,126], and predicts the binding mode and dimerization free energies of membrane proteins in agreement with full atomistic simulations [127-129]. For the soluble parts of membrane proteins, however, there is evidence that the Martini model is too sticky [130]. This might have limited the rearrangements of CIII with its large periplasmic domains. A further limitation is the simplified composition of the mitochondrial membrane in our simulations. In particular, we could not include CI for which only recently the structure was resolved [131]. Experiments have shown the association and function of the respiratory chain complexes to depend on the complexes available in the membrane [132,133]. Therefore, the number and quality of the interfaces obtained in our simulations is without doubt affected by the absence of CI. Keeping these limitations in mind, our data provide fuel for the ongoing discussion on the formation of supercomplexes.

Recently, a more dynamic view is emerging for the organization of the supercomplexes in the respiratory chain [19]. Instead of being purely static, it is now described as a balance between fluid and static models. The fluidic nature would allow adaptation to the physiological conditions, in a manner comparable to the photosystem in thylakoid membranes [20]. The dynamic reorganization of the supercomplexes (continuous exchange of complexes involved in supercomplexes and the pool of free complexes) is assumed to be several orders of magnitude longer than the time scale simulated in this work. In that sense the limited reorientations we see on the time scale of our simulations are biologically realistic. Although recent EM data point to specific conformation for bovine supercomplex [7-10], the novel view on the subject allows the formation of multiple, possibly transient, interfaces both in stoichiometry and orientation in line with our simulations.

The kinetics of lipid exchange and the equilibration of the lipid content at the interfaces between proteins are really fast com-

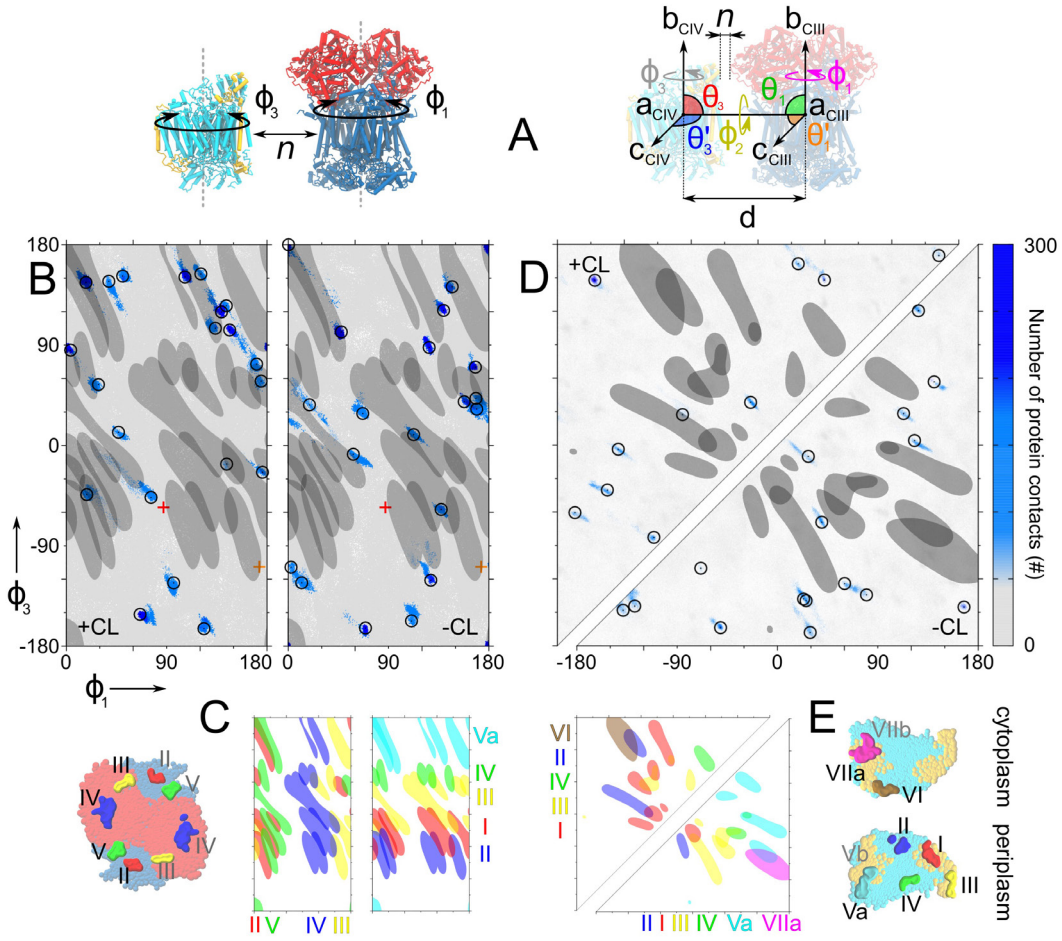

Figure 5.6 | Characterization of protein-protein interfaces. **A)** Schematic representation of the quantities used to report on the relative orientation of the complexes in a supercomplexe. In this example of a CIII/CIV supercomplex, ϕ_1 angle reports on the side CIII (or CIV) exposes to the interface with CIV, while ϕ_3 on the side CIV exposes to CIII. n stands for minimal distance between complexes and/or number of contacts and/or buried surface. **B)** Projection map of CIII and CIV relative orientations in the supercomplexes observed in the simulations for the system with CLs in the membrane bulk on the left and without CLs on the right. The black circles indicate relative orientation of CIII and CIV in the supercomplex formed at the end of the simulations. A grey to dark blue gradient indicates the number of protein contacts formed between the complexes and thus reports on the degree of association of the complexes. The darker the blue the more contacts are formed. The dark grey overlaid zones depict the ranges of ϕ_1/ϕ_3 angles where CL binding sites are shared between complexes; their labels list the site on CIII and then the site on CIV. In the case of the system containing CLs, the supercomplex in formation is not reported here. The red and orange crosses indicate the relative orientation of complexes found in bovine [9] and the yeast [11] supercomplex models, respectively. **C)** Extent of CL binding sites interaction (in terms of ϕ_1/ϕ_3) for a CIII/CIV supercomplex. Same zones colored according to the site involved, for each complex; the complexes with the overlaid binding sites are reported here to ease the visualization. **D)** Relative orientation map of CIV/CIV supercomplexes: the system containing CLs in the lipid bulk on the top triangle, system without CLs on the bottom triangle. Again, the overlaid zones correspond to the ranges of ϕ_1/ϕ_3 angles where CL binding sites are interacting; for each zone, the first roman number represents the site on CIV, the second one the site on CIV. **E)** Extent of CL binding sites interaction (in terms of ϕ_1/ϕ_3) for a CIV/CIV supercomplex.

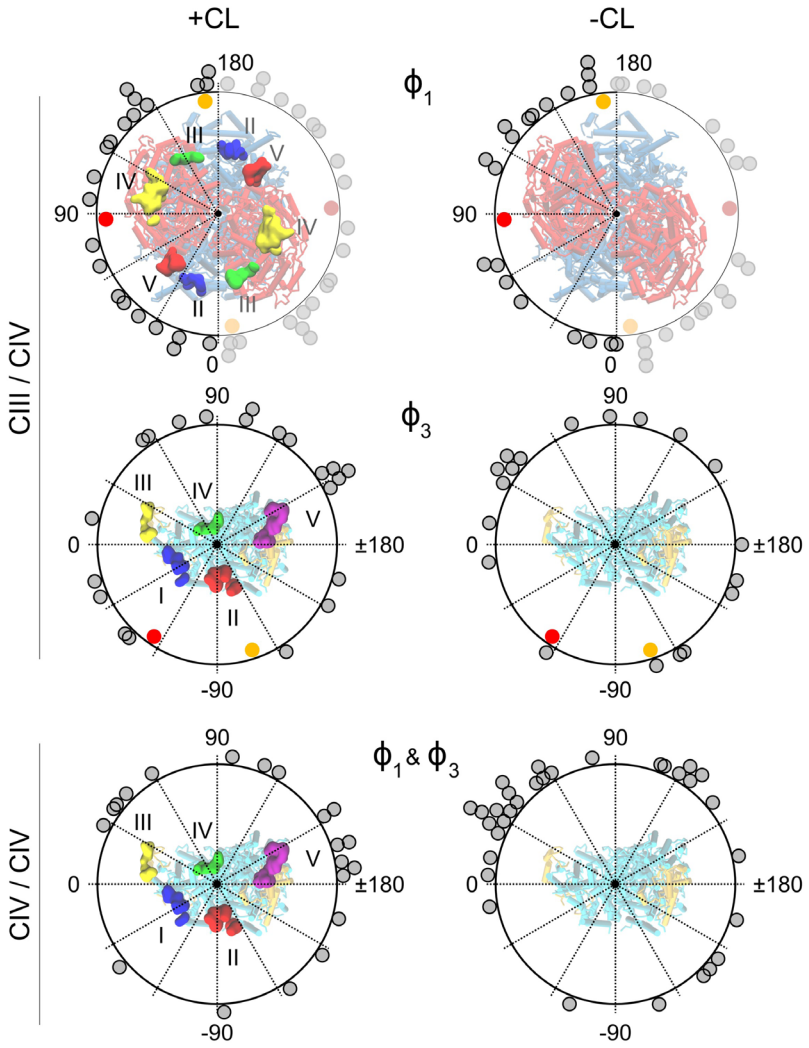

Figure 5.7 | Projection of the interfaces formed on each type of complexes with (+) and without (—) CL. A gray dot represents an interface with ϕ_1 and ϕ_3 reporting on the side of CIII or CIV interacting with CIV or CIII, respectively. The value of the angles is projected onto a circle surrounding the corresponding complex. In the case of CIV-CIV interfaces, the two complexes are not differentiated. The red and orange dots report the relative orientations found experimentally for bovine [9] and yeast [11] experimental models. In the case of the system containing CLs, the supercomplex in formation is not reported here.

pared to the total simulation length, and thus included in our trajectories. The main hypothesis tested in this study is how CLs are "gluing" the respiratory chain complexes together [28], and this dynamics can therefore be assessed. Two propositions have arisen, both consistent with the existence of CL binding sites on the surface of the complexes: these binding sites could i) be directly involved at the interface, bridging the proteins, or ii) be blocking (by occupying) specific interfaces, locking the relative orientation of the complexes away from biologically unfavorable interactions. Our simulations show these binding sites, initially fully populated, keep their CL population constant over the entire simulation regardless of the interactions between complexes. We find CLs to always be present at the interface between complexes, interfaces which are strongly

enriched compared to the bulk concentration (1:4 CL:POPC ratio at the interfaces *vs.* 1:15 in the bulk). Although the results of the relative orientation analysis on our limited statistics (~20 interfaces formed in each system) is showing only a slight preference for interfaces involving CL binding sites, a steering role of CL favoring interactions through site V and disfavoring interactions through site III is evident. Altogether, our data point to: i) a lubricating role of CLs, which can be attributed to their bulky tails. This observation is consistent with the current view available in the literature, suggesting a role of lipids in reorganization of the complexes, allowing reorientation and unbinding [134]; ii) a presence of shared binding sites (bridges), shown for instance in some of the graphical snapshots (Fig. 5.7); and iii) a steering role of CL, favoring interactions through site V and disfavoring site III (Fig. 5.6) on CIV. The latter observation would point to a possible role of lock as well, and to a more complex role of CLs beyond the simplistic either/or view of bridge or lock.

Another point worth discussing is the amount of lipids lubricating the complex interfaces. Recent experimental studies claim for a large distance between complexes and estimate the presence of up to ~50 (yeast CIII/CIV supercomplex [12]) or ~200 (bovine CI/CIII/CIV supercomplex [10]) CLs in the vicinity of the formed supercomplexes. Even though the interfaces obtained in our simulations are apparently much tighter than the model suggested by the electron density maps, a rough count of the lipid population yield to the same numbers: the final value of the plots reported on Figure 5.4 give an average of 35 CLs around CIII and 20 around CIV. From the sum of these values should be substracted the common CLs, *i.e.* the CLs at the interface between complexes and thus counted two times; the number of neighbouring CLs is similar to what was estimated experimentally for a supercomplex CIII/CIV.

In conclusion, we have described here the very first simulations, at near-atomic detail, of the formation of respiratory chain supercomplexes in model mitochondrial membranes. Our data point to a significant role of CL in the process, steering complex association. However the molecular mechanism is not as clearly marked as previously hypothesized, and warrants further studies. In a more general sense, this work opens the way to simulate large-scale protein aggregation in realistic membrane environments. §